Scientometrics (2011) 86:99–112
DOI 10.1007/s11192-010-0205-9

On the map: *Nature* and *Science* editorials

Cathelijn J. F. Waaijer · Cornelis A. van Bochove · Nees Jan van Eck

Received: 9 February 2010 / Published online: 13 March 2010

Abstract Bibliometric mapping of scientific articles based on keywords and technical terms in abstracts is now frequently used to chart scientific fields. In contrast, no significant mapping has been applied to the full texts of non-specialist documents. Editorials in *Nature* and *Science* are such non-specialist documents, reflecting the views of the two most read scientific journals on science, technology and policy issues. We use the VOSviewer mapping software to chart the topics of these editorials. A term map and a document map are constructed and clusters are distinguished in both of them. The validity of the document clustering is verified by a manual analysis of a sample of the editorials. This analysis confirms the homogeneity of the clusters obtained by mapping and augments the latter with further detail. As a result, the analysis provides reliable information on the distribution of the editorials over topics, and on differences between the journals. The most striking difference is that *Nature* devotes more attention to internal science policy issues and *Science* more to the political influence of scientists.

Keywords Bibliometrics · Classification · Editorials · Full-text · Mapping · VOSviewer

Introduction

In bibliometrics, mapping is an increasingly important tool in the classification of documents into groups and subgroups and in the analysis of other types of patterns (e.g., Börner

Electronic supplementary material The online version of this article (doi:10.1007/s11192-010-0205-9) contains supplementary material, which is available to authorized users.

C. J. F. Waaijer · C. A. van Bochove (✉) · N. J. van Eck
Centre for Science and Technology Studies, Leiden University, Leiden, The Netherlands
e-mail: cbochove@cwts.leidenuniv.nl

C. J. F. Waaijer
e-mail: c.j.f.waaijer@lumc.nl

N. J. van Eck
e-mail: ecknjpvan@cwts.leidenuniv.nl

et al. 2003). So far, mapping techniques have mostly been applied to data extracted from scientific documents. The raw materials for bibliometric maps have been citations, keywords and technical terms in titles and abstracts. Mapping has provided information on issues such as relations between scientific fields (e.g., Noyons and Van Raan 1998; Van Eck and Waltman 2007), relations between scholars or journals (e.g., McCain 1991; White and McCain 1998) and scientific collaboration between scholars, institutions or countries (e.g., Luukkonen et al. 1993; Peters and Van Raan 1991).

So far, little or no mapping has been attempted in the analysis of bodies of non-scientific or non-specialist documents. It is not clear whether mapping is a useful tool in this case where citations, keywords and abstracts with technical terms are not available. The raw material for mapping would have to be the words in the full texts of the documents, but it has yet to be seen whether in non-specialist documents the relation between the content and the words used is strong enough to generate meaningful patterns through mapping. However, it would be very important if mapping turns out to be effective, since in that case mapping might to some degree replace the traditional manual analysis of bodies of documents. Here manual refers to determining the content of the documents by actually reading them (perhaps partially or superficially) and classifying them into groups and subgroups on the basis of their content.

In this paper we apply mapping to a body of general, non-specialist documents. The body of documents concerned is the editorials of *Nature* and *Science* from 2000 on (Waaijer et al. 2010). These documents are important and interesting in their own right, because they reflect the views of the two most read scientific journals on what topics are important in the conduct and application of scientific research. Differences between *Nature* and *Science* might be interesting, as they could reflect differences between European and US views or differences in perspective between the editors of an independent commercial publisher (*Nature*) and the editors of a learned society journal (*Science*).

In view of the novelty of the application of mapping to non-specialist documents, we combined the mapping with a manual classification procedure for a large sample of editorials. We used this method to validate the mapping, notably the interpretation and homogeneity of clusters. In addition, manual classification made it possible to augment our results with supplementary information that is useful in the analysis of differences between *Nature* and *Science*.

The remainder of this paper is organized as follows. The methods that were used to map and classify the editorials will first be expanded upon. Subject areas and their relations in the editorials will then be shown by constructing two maps, a term map and a document map. Subsequently, the combination of bibliometric mapping and manual classification will be shown in the validation of the document map. Finally, the validated document map will be used to point out differences between *Nature* and *Science* and between an early and a late period.

Data and methods

Data collection

In the first step of the data collection process, all *Nature* and *Science* editorials published between 1 January 2000 and 2 July 2009 were retrieved in HTML format. In total, 1,565 editorials were retrieved, 1,097 from *Nature* and 468 from *Science*. After retrieving the editorials, the full text of the editorials was extracted from the HTML files.

Term identification

Two maps were constructed, a term map and a document map. To construct these maps, terms needed to be identified in the editorials. Since manual term identification is subjective and labour intensive, we took an automatic term identification approach. We first used computer programme NPtool (Voutilainen 1993) to identify noun phrases in the editorials. Most noun phrases were identified correctly using this programme. However, noun phrases containing a conjunction or preposition, such as 'Food and Drug Administration' and 'National Academy of Sciences', were not identified correctly. To solve this problem, we created a lexicon of noun phrases containing a conjunction or preposition. Using this lexicon, these noun phrases could be identified correctly. The criterion for a noun phrase to be included in the lexicon was that a fragment of the noun phrase (e.g., 'Drug Administration') occurs at least five times in the editorials and that the complete noun phrase (e.g., 'Food and Drug Administration') appears on the first page of the Google search engine when searching for the fragment. The lexicon can be found in the Supplementary Material (p. 66). After identifying the noun phrases, we calculated for each noun phrase its so-called termhood. This is a measure that indicates to what degree a noun phrase is systematically associated with specific underlying topics (Van Eck et al. 2010). Of all noun phrases occurring at least 15 times in the editorials, the 600 noun phrases with the highest termhood were selected to be used in the construction of the term and document maps. In the rest of this paper, we refer to these noun phrases as terms.

Term map construction

A term map is a map that shows the relations between terms in a certain domain. In general, the closer two terms are located to each other in a term map, the stronger the relation between the terms. Term maps are also referred to as co-word maps (e.g., Peters and Van Raan 1993).

A term map of the 600 terms identified in the *Nature* and *Science* editorials was constructed as follows. For each pair of terms, the number of co-occurrences was counted. The number of co-occurrences of two terms is the number of times that they occur jointly in an editorial. If a term occurs more than once in an editorial, this yields more than one co-occurrence with the other terms in that editorial (e.g., if terms X and Y occur, respectively, two and three times in a single editorial, this yields six co-occurrences). Based on the co-occurrence counts, the similarity of terms was calculated using the association strength measure discussed by Van Eck and Waltman (2009). The similarities were used as input for the VOS mapping technique (Van Eck et al. submitted). Based on the similarities, the VOS mapping technique determined a location in a two-dimensional map for each of the 600 terms. The objective of the VOS mapping technique is to locate terms with a high similarity close to each other and terms with a low similarity far away from each other. However, since only two dimensions are available, this objective usually cannot be achieved perfectly. The VOS mapping technique then attempts to approximate the objective as closely as possible. The VOS mapping technique can be seen as an alternative to the well-known technique of multidimensional scaling. An in-depth comparison of the two techniques is provided by Van Eck et al. (submitted). The comparison shows that in general the VOS technique provides more satisfactory representations of data sets than the multidimensional scaling technique. A computer programme called VOSviewer (Van Eck and Waltman in press) was used to visualize the map produced by the VOS mapping technique.

As a further step in the analysis, the 600 terms identified in the *Nature* and *Science* editorials were assigned to clusters. This was done using a clustering technique that relies on a multinomial mixture model (similar to Zhu et al. 2009, Section 2.3). The assignment of terms to clusters was based on the editorials in which a term occurs. Six clusters were used, since this number seemed to yield the most easily interpretable results. The VOSviewer software was used to visualize the assignment of terms to clusters.

Document map construction

A document map is a map that shows the relations within a set of documents (e.g., Åström 2007; Janssens et al. 2006; Klavans and Boyack 2006). In general, the closer two documents are located to each other in a document map, the stronger the relation between the documents.

A document map of the 1,565 *Nature* and *Science* editorials was constructed in a similar way as the term map discussed above. For each pair of editorials, the number of co-occurrences was counted. The number of co-occurrences of two editorials is the number of terms that occur in both editorials. Again, terms that occur more than once in the same editorial can yield more than one co-occurrence. After counting co-occurrences, similarities were calculated using the association strength measure and the VOS mapping technique was applied to the similarities. The VOS mapping technique determined for each of the 1,565 editorials a location in a two-dimensional map. The VOSviewer software was used to visualize the map produced by the VOS mapping technique. For each editorial, additional information such as the title, the text of the first paragraph and a list of important terms was also provided to the VOSviewer software. This information served to simplify the interpretation of the map.

The document map as such is useful to explore what topics the editorials are about and how the topics are related. However, a quantitative analysis of the topics requires that the editorials are grouped into clusters that are associated with topics. The clustering technique that was used is different from the one used to cluster the terms. To cluster the editorials, the well-known K-means algorithm was applied to the coordinates of the editorials in the document map. Because a reasonably fine-grained clustering was needed, it was decided to use 15 clusters. The VOSviewer software was used to visualize the clustering of the editorials. Since the clustering is based on the document map, we will refer to it as a map-based clustering later on in this paper. Note that the clusters have been determined by a statistical technique and not by an a priori delineation of topics. Naturally, it is to be hoped that the clustering technique leads to recognizable topics, but it has to be explicitly investigated whether this is actually the case.

Validation and content-based analysis of document map

To determine whether the document clusters refer to recognizable topics, the content of each cluster needs to be identified and an appropriate label must be assigned to capture the essence of the content. This requires an iterative process of analysis and interpretation.

The first step of this process is to inspect a number of elements from each cluster and to give a characterization of these elements. This characterization must both be intuitively comprehensible and 'predictive' of the characteristics of other elements from the same cluster. Next, some of these other elements are studied to verify whether they fit the 'predictions'. If this turns out to be the case, the cluster can be considered homogeneous with respect to the characterization and can be assigned a label.

However, if the predictions are not borne out by the newly inspected elements, the characterization of the cluster needs to be adjusted and the process is repeated. The adjustment may be a modification that corrects for errors in the original characterization, but more often it amounts to a generalization that makes the characterization applicable to more elements. Naturally, this generalization comes at a price. It causes distinctions between clusters to become less sharp and characterizations to overlap. Therefore, it may be necessary to sharpen the characterization again, implying that some of the elements of the cluster do not fit the characterization. Thus, the iterative process essentially searches for characterizations that balance on the one hand the amount of overlap and on the other hand the number of cluster elements that do not fit the characterization. If no reasonable balance can be found for a considerable part of the clusters, the whole clustering needs to be rejected and a new approach (e.g., changing the number of clusters or changing the terms used for the mapping exercise) has to be adopted.

In a term map the characterization problem is somewhat easier to solve than in a document map. As the elements of the clusters are terms, a characterization amounts to providing a general heading for the terms in a cluster and inspecting whether a considerable majority of the terms in a cluster do indeed fall under this heading. However, in the case of a document map, characterization may be quite difficult and at the same time require a high degree of accuracy. Occasionally, the titles of the documents may help, but in the case of editorials these often are intended as a pun, phrased to capture attention and not very informative on the subject matter. Therefore, the key terms in an editorial are the most important information to work with. This is why it is important that in the VOSviewer software one can zoom in on individual editorials and view not just their title and first paragraph but also the terms that are most specific for the editorial.

Using the mapping and visualization technique, the content of most clusters could be determined fairly well, but the proper characterization of some clusters remained somewhat uncertain or elusive. This is unacceptable if, as in the present case, a high degree of accuracy is required.

For this reason, we employed a powerful validation method for the characterization of the clusters. We read and summarized a sample of editorials from each cluster (Supplementary Material Table 1). At least 10 editorials from *Nature* and 10 from *Science* were read from each cluster. Using our summaries, the sample editorials from each cluster were classified into subgroups with a homogeneous content. In most clusters, most sample editorials were immediately seen to be part of one or a few homogeneous subgroups. From these subgroups, the main content of the cluster could then be determined quite accurately and a complete content classification of the editorials could be drawn up (Supplementary Material Table 2).

The content of some editorials in some clusters did not fit into the subgroups belonging to the cluster but instead fitted into another cluster. A very small number of editorials actually did not fit into any of the clusters at all. Consequently, each sample editorial now has two classifications:

(1) The map-based cluster to which the editorial was assigned by the clustering technique described above.
(2) The content-based cluster to which the editorial belongs according to the manual classification.

Thus the sample makes it possible to *confirm* the homogeneity of the clusters, to *interpret* the clusters by providing them with an appropriate label and to *augment* the clustering, both by adding detail and by indicating the level of accuracy.

Results

Term map

The term map is shown in Fig. 1. The map can be examined in full detail using the VOSviewer software at www.vosviewer.com/editorials/terms.php. Terms that are located close to each other in the map often occur together in the same editorial, while terms that are located far away from each other do not or almost not occur together. In general, terms in the centre of the map co-occur with many different terms and are therefore related to various topics. In contrast, terms at the edges of the map tend to co-occur only with a small number of other terms. Terms at the edges therefore often belong to relatively isolated fields. The colour of a term indicates the cluster to which the term has been assigned, and the size of a term indicates the frequency with which the term occurs in the editorials. The size of a cluster in the map is influenced by many factors (e.g., the number of terms in the cluster, the frequency of occurrence of the terms and the strength with which the terms are related to each other) and therefore does not have a straightforward interpretation. The density of an area in the map is determined by the number of terms in the area and by the frequency with which the terms occur in the editorials.

Clusters that can be easily interpreted in the term map are space and physics (in red), the scientific publication system (in yellow), stem cell research (in light blue), bioscience (in

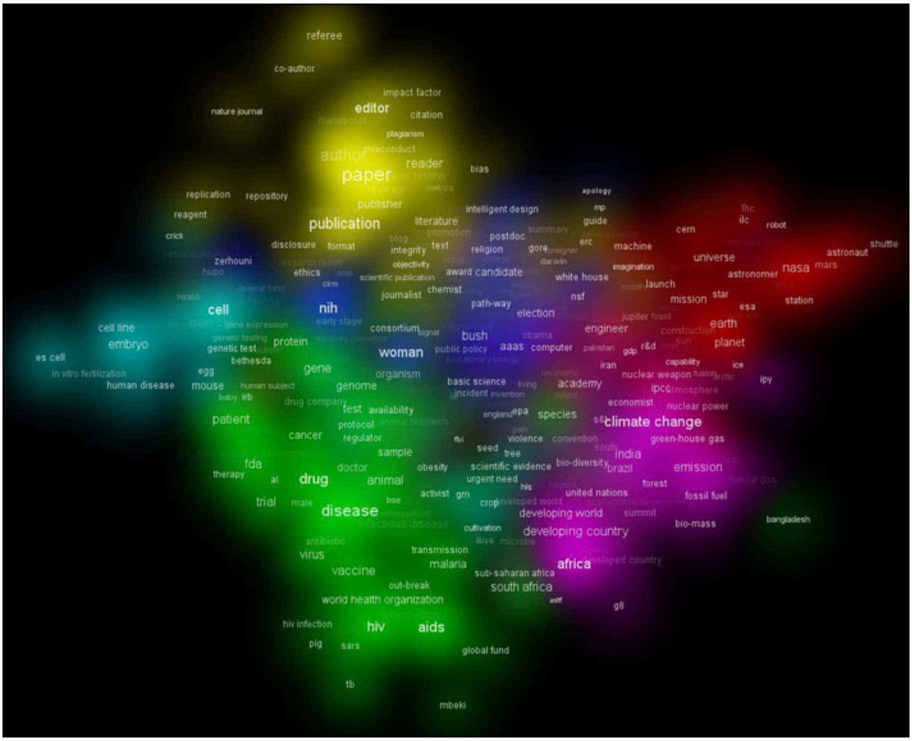

Fig. 1 Term map of the 600 terms identified in the *Nature* and *Science* editorials. At www.vosviewer.com/editorials/terms.php the map can be examined in full detail using the VOSviewer software

green) and global problems (in pink). There is a more poorly defined cluster in dark blue, with terms related to politics. This cluster is located more or less in the centre of the map, which shows that politics is related to many different topics. In contrast, the space and physics cluster (in red) and the stem cell cluster (in light blue) are located more towards the edges of the map, which suggests that these topics are somewhat unrelated to other topics.

The size of the terms and the density of the different areas indicate that the scientific publication system receives much attention. The core terms are ‘paper’, ‘author’, ‘publication’ and ‘editor’. Other terms in this cluster suggest that it deals with the way papers are published (‘peer review’, ‘reviewer’, ‘submission’, ‘repository’), with bibliometrics (‘impact factor’, ‘citation’, ‘metrics’) and with scientific integrity (‘plagiarism’, ‘misconduct’, ‘research misconduct’, ‘scientific misconduct’, ‘validity’, ‘replication’, ‘integrity’, ‘ethics’).

Although the clusters on space and physics and on stem cell research are located more towards the edges of the map, they do have locations close to areas one would expect them to be related to. In case of the stem cell cluster, nearby terms are related to the ethical issues of drug trials, such as ‘IRB’, ‘human subject’ and ‘patient’. The same applies to terms related to genetic testing. In case of the space and physics cluster, both the politics cluster and the global problems (especially climate change) cluster are nearby. This is due to the fact that space and physics research requires a large amount of funding from governmental organizations and, in case of global problems, to the fact that terms such as ‘earth’ and ‘planet’ occur both in space and physics and in climate change editorials.

The term map shows a contingency between terms such as HIV/AIDS and other infectious diseases on the one hand and developing countries on the other hand. This indicates that *Nature* and *Science* mainly write about infectious and neglected diseases in relation to developing countries. Similarly, terms concerning developing countries are in the same region of the map and in the same cluster (we already referred to this cluster as ‘global problems’) as terms concerning climate change. This suggests that quite a large number of editorials deal with the relation between climate change and developing countries.

Terms concerning education (‘teaching’, ‘classroom’, ‘teacher’) are located close to ‘religion’ and ‘intelligent design’, in the dark blue cluster. This cluster also contains a considerable number of terms from politics. Clearly, editorials of *Nature* and *Science* pay serious attention to the politics of religion and evolution in the classroom.

Document map

A term map of the main terms in a corpus of documents gives a good overview of the subject areas in the corpus. However, it only shows the relations between the terms in the documents, not necessarily the relations between the documents themselves. We are interested in investigating possible differences in topic choice between *Nature* and *Science* and between an early and a late period. Therefore, we constructed a document map and identified 15 clusters based on the locations of the editorials in the document map. A first iterative analysis of this map with the VOSviewer gave the impression that the clustering of editorials is good at the edges of the map but that the clusters in the centre of the map might be less coherent.

As announced in the methods section, to confirm the homogeneity of the clusters in the document map and to aid in their interpretation, we read, summarized and classified a sample of editorials. We used a sample design with sufficient resolution to determine

Fig. 2 Document map of the 1,565 *Nature* and *Science* editorials after content-based labelling of the clusters. At www.vosviewer.com/editorials/editorials.php the map can be examined in full detail using the VOSviewer software

differences between *Nature* and *Science* in the amount of attention to the various topics. The results were used to establish labels that best characterize the content of each of the clusters in the document map. The document map together with the cluster labels is shown in Fig. 2. At www.vosviewer.com/editorials/editorials.php the map can be examined in full detail using the VOSviewer software.

The 15 clusters of editorials are listed in Table 1. The clusters can be aggregated into five groups that roughly correspond to the topics identified in the term map: the scientific publication system (journal policies, science publication), biomedical issues (biopolicies, bioscience, drug development, infectious diseases and toxins, NIH, health), generalized science policy (science policy, research climate, science organization, science and society), global problems (climate change, developing countries and global problems), and space and physics.

The 'goodness of fit' of the map-based clustering was assessed first by comparing the distribution of the editorials over the map-based clusters with the distribution of the editorials over the content-based clusters. These two distributions are reported in Table 1 (for a more detailed analysis, see Supplementary Material Table 3). The first distribution is based on the map-based clustering. The second distribution is based on the manual classification of the sample of editorials. The sample results have been raised to population totals using the inverse of the sample fraction. This has been done to achieve easy comparability with the population-based results from the map-based clustering.

Table 1 shows that the map-based and content-based distributions differ only marginally. The most important differences are in the science policy clusters. About 10% of the editorials belong to the two publication clusters. Of this 10%, 4% is about the rules and products of *Nature* and *Science* themselves and 6% is about more general issues of scientific publishing. Almost 30% of the editorials have been assigned to the six biomedical clusters, 40% belongs to the four science policy clusters, close to 20% to the two clusters on global problems and 5% to the space and physics cluster.

Table 1 compares the *balance* of the map-based and content-based distributions. The effect of an editorial belonging to map-based cluster X and content-based cluster Y is cancelled out by the effect of an editorial belonging to map-based cluster Y and content-

Table 1 Map-based and content-based percentage distributions of the editorials

	Map	Content		Map	Content
Publication system	10	9	Generalized science policy	39	39
Journal policies	4	4	Science policy	12	16
Science publication	6	4	Research climate	5	5
Biomedical issues	29	27	Science organization	13	11
Biopolicies	7	6	Science and society	8	7
Bioscience	4	4	Global problems	18	18
Drug development	4	3	Climate change	9	10
Infect. diseases, toxins	7	7	Dev. countries, global problems	9	8
NIH	2	2	Space and physics	5	5
Health	6	5	Space and physics	5	5

Table 2 Transition table showing for each map-based cluster the percentage distribution of editorials over the content-based clusters

		Content-based cluster																Total
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Map-based cluster																		
1	Science policy	90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	100
2	Journal policies	6	88	0	0	0	0	3	0	0	0	0	0	3	0	0	0	100
3	Drug development	0	6	94	0	0	0	0	0	0	0	0	0	0	0	0	0	100
4	Space and physics	0	0	0	94	0	0	0	0	6	0	0	0	0	0	0	0	100
5	Bioscience	3	0	0	0	85	0	0	0	0	6	0	0	0	0	0	6	100
6	Biopolicies	0	0	0	0	0	82	0	0	0	7	0	3	0	3	5	0	100
7	Research climate	0	0	0	0	0	7	75	0	10	0	0	7	0	0	0	0	100
8	Health	0	0	0	0	0	9	0	82	0	9	0	0	0	0	0	0	100
9	Climate change	7	0	0	0	0	0	0	0	93	0	0	0	0	0	0	0	100
10	Science organization	18	0	0	0	0	0	8	0	2	63	0	0	0	0	0	10	100
11	NIH	0	0	0	0	0	0	0	0	0	3	97	0	0	0	0	0	100
12	Science and society	15	0	0	0	0	0	0	0	0	4	0	76	3	0	0	3	100
13	Science publication	3	9	0	0	0	0	0	0	0	13	0	6	69	0	0	0	100
14	Dev. countries, global problems	0	0	0	0	0	0	4	0	0	6	0	0	0	90	0	0	100
15	Infect. diseases, toxins	0	0	0	0	0	0	0	0	0	0	0	3	0	0	97	0	100

based cluster X. A full comparison of the map-based and content-based clusterings can be made using a *transition table*. Table 2 provides a transition table showing for each map-based cluster the distribution of editorials over the content-based clusters. This is the most informative transition table for the validation of the map-based clustering. However, in the Supplementary Material we also provide a transition table showing for each content-based cluster the distribution of editorials over the map-based clusters (Supplementary Material Table 4). Furthermore, transition tables can be constructed for *Nature* and *Science* separately. Such tables are also provided in the Supplementary Material. It turns out that the transition patterns for the two journals are quite similar.

The main diagonal of Table 2 indicates for each map-based cluster the percentage of editorials that have been assigned correctly. This provides a direct verification of the quality of the map-based clustering. In close to half of the clusters at least 90% of the editorials is on the main diagonal, and in all but one of the clusters at least two-third of the editorials is. Our first impression that the map-based clustering is more accurate at the edges of the map than in the centre is borne out by the transition table. All clusters at the edges of the map (space and physics, climate change, developing countries and global problems, biopolicies, infectious diseases and toxins, drug development, health, NIH, journal policies) have main diagonal values of at least 80%, while most of the clusters in the centre of the map (science organization, science and society, research climate, science publication) have lower main diagonal values. However, the central clusters on science policy and bioscience are exceptions to the rule, since their main diagonal values are quite high.

Put succinctly, the sample-based content classification essentially *confirms* the results of the map-based clustering. Perhaps the most important contribution of the manual classification is to clarify the interpretation of the map-based clusters, that is, to characterize the content of the clusters and to provide appropriate labels.

Topic choice differences between *Nature* and *Science*

The map-based clustering of editorials made it possible to investigate differences in topic choice between different subsets of editorials. We first investigated whether there are any differences between editorials published in *Nature* and editorials published in *Science*. For each map-based cluster, we calculated the percentage of editorials published in each of the two journals. The percentages were normalized for the fact that the total number of *Nature* editorials in the entire corpus is more than twice as high as the total number of *Science* editorials.

In Fig. 3 the differences between *Nature* and *Science* are shown using pie charts. On the whole, the distribution of editorials over the 15 clusters is quite similar for *Nature* and *Science*. However, there are some intriguing differences. The largest difference is in space and physics. *Nature* devotes three times more editorials to this topic (including a substantial number of editorials on NASA) than *Science* (6% vs. 2%). This was confirmed by the content analysis.

A remarkable difference concerns the NIH. Two percent of the *Nature* editorials are on this US agency and just 1% of the *Science* editorials. This was again confirmed by the content analysis. In fact, about half of the *Nature* editorials in the NIH cluster concern the organization and management of NIH. One of the editorials points to the reason why the European *Nature* writes so much about this US medical research agency: NIH is the largest research agency in the world. In fact, it is more remarkable *Science* writes so little about NIH.

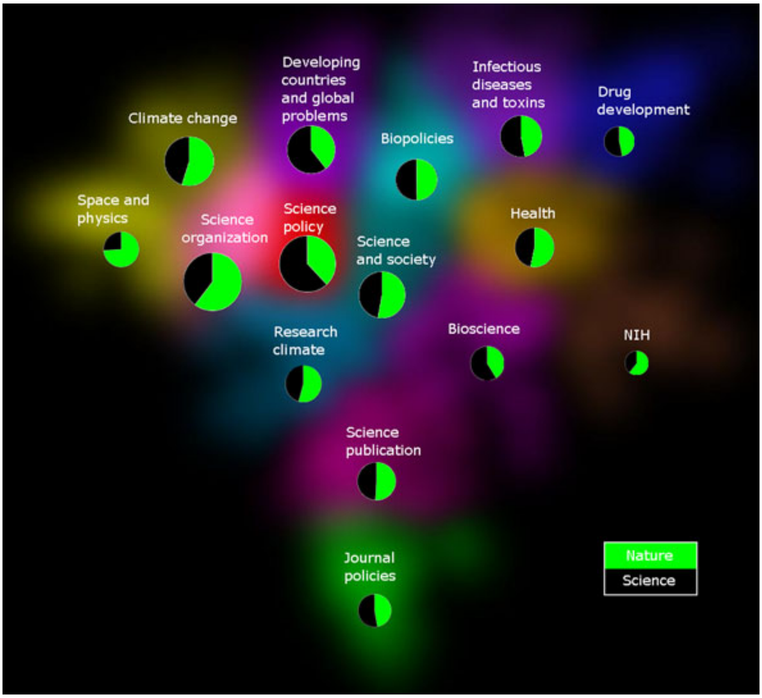

Fig. 3 Differences in topic choice between *Nature* and *Science* depicted in the document map. The size of a pie chart indicates the number of editorials in the corresponding cluster

A further significant difference between *Nature* and *Science* is that a larger percentage of the *Science* editorials is about developing countries, environmental protection, climate change and other global problems. This is mainly due to the cluster on developing countries and global problems. In contrast, *Nature* and *Science* devote approximately the same amount of attention to the related cluster on climate change (10% of the editorials).

Perhaps the most striking difference between *Nature* and *Science* concerns the science policy clusters. At first glance, looking at the map-based clustering only, *Science* writes more than *Nature* (17% vs. 10%) about science policy in a narrow sense (policies to maximize scientific output, such as priority setting, research quality management and impact of science on political decision making) and less about science organization issues. Looking at the content-based clustering, this difference is almost eliminated. Fifteen percent of the *Nature* editorials is then seen to deal with science policy. Things become really intriguing if we look within this content based cluster. *Nature* turns out to devote more attention to priority setting, while *Science* is more interested in the political influence of science and scientists. Moreover, a number of editorials of both *Nature* and *Science* belonging to one of the biomedical clusters or to the space and physics cluster also deal with priority setting in these fields. Taking this into account, almost 15% of the *Nature* editorials deals with priority setting, whereas only 8% of the *Science* editorials does. It appears that *Science* is more reticent than *Nature* in dealing with sensitive within-science issues. This would merit a study into the question whether this difference in editorial policy can be attributed to the greater independence of the commercially published *Nature* from the scientific establishment.

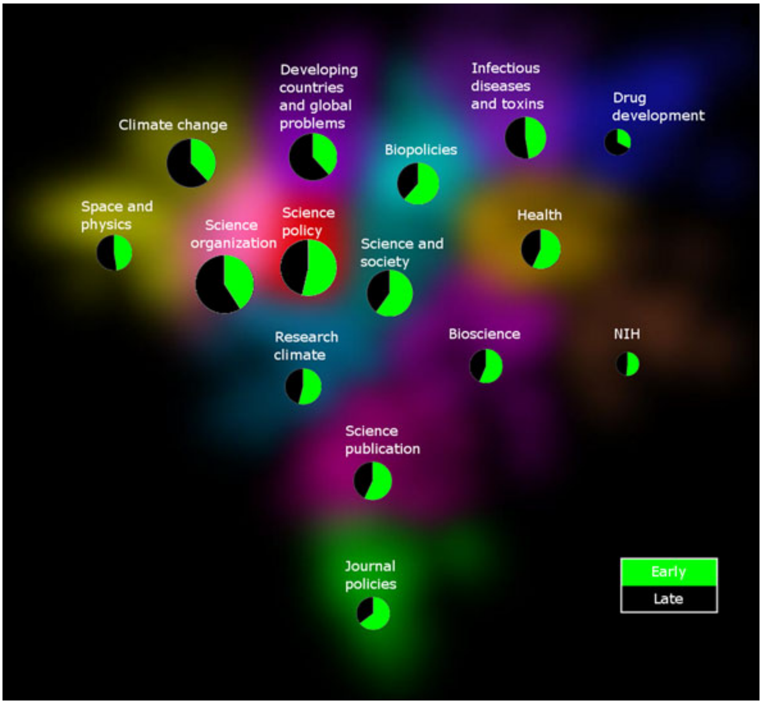

Fig. 4 Differences in topic choice between an early (2000–mid 2004) and a late (mid 2004–mid 2009) period depicted in the document map. The size of a pie chart indicates the number of editorials in the corresponding cluster

Differences in topic choice between an early and late period

In addition to differences between *Nature* and *Science*, we also investigated possible differences over time. For this purpose we divided the entire period into an early (2000–mid 2004) and a late (mid 2004–mid 2009) period. A normalization was applied for differences in the total number of editorials in each of the two periods. As can be seen in Fig. 4, in the late period there was more attention for developing countries and global problems, drug development and climate change. This reflects the increased attention for climate change during the past years. Conversely, in the early period *Nature* and *Science* devoted more attention to journal policies, science and society issues and biopolicies.

Conclusion

In this paper we have shown that it is possible to classify a body of full-text, non-specialist documents using a two-step method that combines bibliometric mapping techniques and manual classification. Our analysis was performed on the editorials of *Nature* and *Science* published between 2000 and mid 2009. The words used in these editorials are less specialist than the words used in titles and abstracts of scientific papers, which are more commonly analysed using bibliometric mapping techniques. In addition, editorials contain between 500 and 1,000 words, which is much more than the average number of words in abstracts of scientific papers.

We used a combination of bibliometric mapping techniques and manual classification of a sample of editorials. The manual classification largely *confirmed* the mapping results. In addition, the manual classification also allowed for a better *interpretation* of the mapping results. Furthermore, the manual classification *augmented* the mapping results with additional details, in particular a further breakdown of the clusters into subgroups.

These findings suggest the recommendation to apply bibliometric mapping techniques to bodies of documents in combination with a manual analysis of a sample of documents, for the purpose of confirmation, interpretation and augmentation. The stratification of the sample using map-based clusters allows a high resolution with a modest absolute sample size.

Acknowledgments We would like to thank Renald Buter for processing the HTML files, for help with the practical analysis and for helpful discussions. In addition, we would like to thank Ed Noyons, Ludo Waltman and Ton van Raan for critical discussions and for offering ideas during the course of this study.

References

Åström, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology, 58*(7), 947–957.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology, 37*, 179–255.

Janssens, F., Leta, J., Glänzel, W., & De Moor, B. (2006). Towards mapping library and information science. *Information Processing and Management, 42*(6), 1614–1642.

Klavans, R., & Boyack, K. W. (2006). Quantitative evaluation of large maps of science. *Scientometrics, 68*(3), 475–499.

Luukkonen, T., Tijssen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics, 28*(1), 15–36.

McCain, K. W. (1991). Mapping economics through the journal literature: An experiment in journal cocitation analysis. *Journal of the American Society for Information Science, 42*(4), 290–296.

Noyons, E. C. M., & Van Raan, A. F. J. (1998). Monitoring scientific developments from a dynamic perspective. Self-organized structuring to map neural network research. *Journal of the American Society for Information Science, 49*(1), 68–81.

Peters, H. P. F., & Van Raan, A. F. J. (1991). Structuring scientific activities by co-author analysis. *Scientometrics, 20*(1), 235–255.

Peters, H. P. F., & Van Raan, A. F. J. (1993). Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Research Policy, 22*(1), 23–45.

Van Eck, N. J., & Waltman, L. (2007). Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, 15*(5), 625–645.

Van Eck, N. J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology, 60*(8), 1635–1651.

Van Eck, N. J., & Waltman, L. (in press). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*.

Van Eck, N. J., Waltman, L., Noyons, E. C. M., & Buter, R. K. (2010). Automatic term identification for bibliometric mapping. *Scientometrics*, *82*(3), 581–596.

Van Eck, N. J., Waltman, L., Dekker, R., & Van den Berg, J. (submitted). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS.

Voutilainen, A. (1993). NPtool, a detector of English noun phrases. In *Proceedings of the Workshop on Very Large Corpora: Academic and Industrial Perspectives*. Ohio State University, Columbus, OH, USA (pp. 48–57).

Waaijer, C. J. F., Van Bochove, C. A., & Van Eck, N. J. (2010). Journal editorials give indication of driving science issues. *Nature, 463*(7278), 157.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327–355.

Zhu, S., Takigawa, I., Zeng, J., & Mamitsuka, H. (2009). Field independent probabilistic model for clustering multi-field documents. *Information Processing and Management, 45*(5), 555–570.